

BPAGS User's Manual

(Version 1.0)

BPAGS Documentation (last updated: January 13, 2024)

Washington State University, Richland, WA 99354, USA

For more information, please contact Suraiya Akhter (Email: suraiya.akhter@wsu.edu)

Please follow the user manual for using our web application. If you find our web application useful, please cite our following paper.

Citation:

Akhter, S. and Miller, J.H., BPAGS: A web application for bacteriocin prediction via feature evaluation using alternating decision tree, genetic algorithm, and linear support vector classifier. *Frontiers in Bioinformatics*, 3, p.1284705.

BPAGS will automatically generate the required features of the testing protein sequences. Furthermore, the web application has the option to use our previously developed BaPreS predictive tool (t-test with SVM) to compare prediction results for the testing sequences. Users can test multiple sequences simultaneously and augment new sequences to the training data to boost the predictive ability of the machine learning models through the web application. The flowchart of our web application is given below.

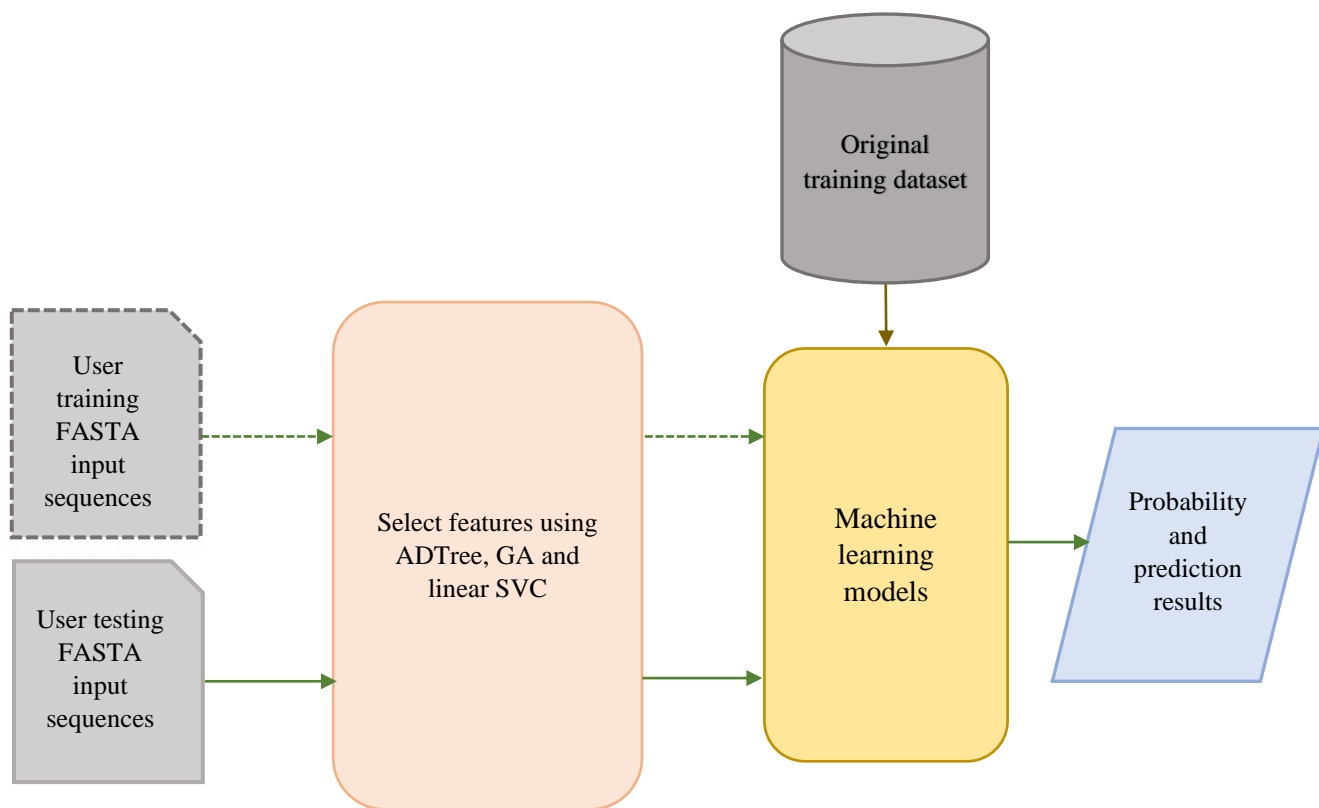


Fig 1: BPAGS workflow.

Home Page:

Fig 2. shows the home page of our web application.

BPAGS: A Web Application for Bacteriocin Prediction

Welcome [Predicted Bacteriocin Sequences](#) [Probability Scores](#) [Help](#) [Data](#)

Upload Files

This web application is used to predict bacteriocin sequences from an input FASTA file. The user manual is provided in the 'Help' menu. To predict bacteriocin sequences, please select a method and upload a FASTA file with the sequences you want to predict using the 'Choose an input fasta file' upload box, following the file format shown at the bottom of this page.

If you wish to add new data to the model (training set), please use the 'Add new bacteriocin sequences to training (FASTA)' upload box for bacteriocin sequences and the 'Add new non-bacteriocin sequences to training (FASTA)' upload box for non-bacteriocin sequences, respectively. Once the necessary files are uploaded, click the 'Bacteriocin prediction' button. This will automatically navigate you to the 'Predicted Bacteriocin Sequences' page once the classification results are generated. Then, click the 'Probability estimation' button to navigate to the 'Probability Scores' page containing the probability values of the sequences.

Download Results

The prediction and probability results can be downloaded by clicking the 'Prediction results' and 'Probability results' buttons, respectively. Additionally, positive and negative datasets, along with the training/testing datasets for all methods, are available in the 'Data' menu.

FASTA Formatting

An example FASTA file can be obtained by clicking on the 'Download Input Samples' button. To predict new sequences, the example FASTA file should be in the form shown below:

[Download Input Samples](#)

```
>BAC110
GGAPATSANAAGAAAVGALAGIPGGPLGVVVGVSAGLTTAIGSTVSGSASSAGGGS
>WP_007226314.1
MSSISKTVAKTGFLILFFISIANAEITGIVSVTDGDTIKVLDENSNQHVKRLTGIDAPE
RQQPFGQASKKYLASMVSGKEVFVESNKKDRYGRVLGKVVVQPADCPSCGKTLDINHAQL
```

Fig 2. home page

Bacteriocin Detection:

To identify bacteriocin and non-bacteriocin sequences, Users need to follow below steps:

1. First, users should select a feature selection method. For instance, in Fig. 3, it is shown that when users choose the ADTree method for feature selection.

BPAGS: A Web Application for Bacteriocin Prediction

Welcome [Predicted Bacteriocin Sequences](#) [Probability Scores](#) [Help](#) [Data](#)

Upload Files

This web application is used to predict bacteriocin sequences from an input FASTA file. The user manual is provided in the 'Help' menu. To predict bacteriocin sequences, please select a method and upload a FASTA file with the sequences you want to predict using the 'Choose an input fasta file' upload box, following the file format shown at the bottom of this page.

If you wish to add new data to the model (training set), please use the 'Add new bacteriocin sequences to training (FASTA)' upload box for bacteriocin sequences and the 'Add new non-bacteriocin sequences to training (FASTA)' upload box for non-bacteriocin sequences, respectively. Once the necessary files are uploaded, click the 'Bacteriocin prediction' button. This will automatically navigate you to the 'Predicted Bacteriocin Sequences' page once the classification results are generated. Then, click the 'Probability estimation' button to navigate to the 'Probability Scores' page containing the probability values of the sequences.

Download Results

The prediction and probability results can be downloaded by clicking the 'Prediction results' and 'Probability results' buttons, respectively. Additionally, positive and negative datasets, along with the training/testing datasets for all methods, are available in the 'Data' menu.

FASTA Formatting

An example FASTA file can be obtained by clicking on the 'Download Input Samples' button. To predict new sequences, the example FASTA file should be in the form shown below:

[Download Input Samples](#)

```
>BAC110
GGAPATSANAAGAAAVGALAGIPGGPLGVVVGVSAGLTTAIGSTVSGSASSAGGGS
>WP_007226314.1
MSSISKTVAKTGFLILFFISIANAEITGIVSVTDGDTIKVLDENSNQHVKRLTGIDAPE
RQQPFGQASKKYLASMVSGKEVFVESNKKDRYGRVLGKVVVQPADCPSCGKTLDINHAQL
```

Fig 3: Method selection

- Next, users can upload the input test data file (fasta file) by clicking on the Browse button under “Choose an input fasta file.” For example, in Fig. 4, the "input_sample.fasta" file is uploaded. During the upload of the input fasta file, all action buttons (Bacteriocin prediction, Probability estimation, Prediction results, and Probability result) are inactive. After completing the upload of data, they will become active and ready to use.

BPAGS: A Web Application for Bacteriocin Prediction

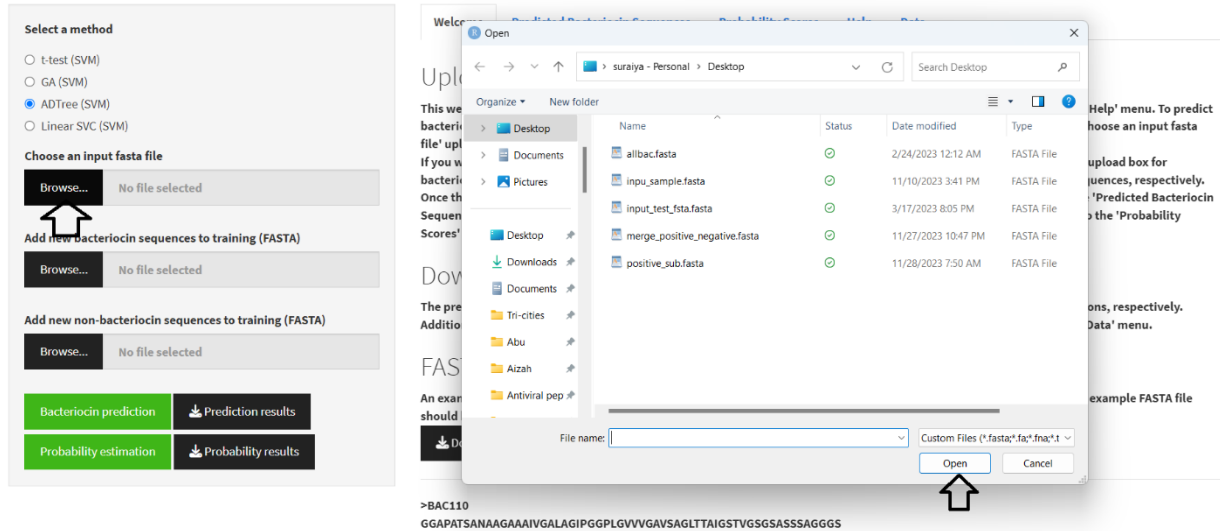


Fig 4: File upload

- Once the necessary files are uploaded, please click the 'Bacteriocin prediction' button first. This will automatically navigate you to the 'Predicted Bacteriocin Sequences' page once the classification results are generated. Refer to the example in Fig. 5 below.

BPAGS: A Web Application for Bacteriocin Prediction

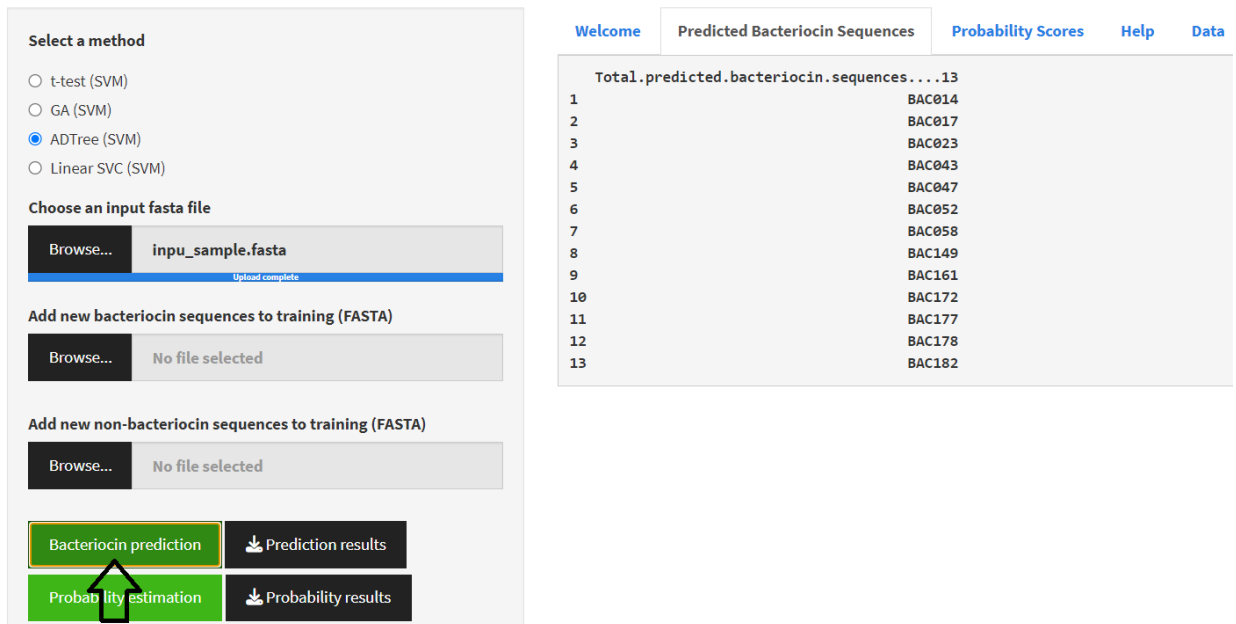


Fig 5: Bacteriocin prediction

- Next, click the 'Probability estimation' button to navigate to the 'Probability Scores' page containing the probability values of the sequences. Refer to the example in Fig. 6 below.

BPAGS: A Web Application for Bacteriocin Prediction

The screenshot shows the BPAGS web application interface. On the left is a sidebar with the following sections:

- Select a method**: Radio buttons for t-test (SVM), GA (SVM), **ADTree (SVM)** (selected), and Linear SVC (SVM).
- Choose an input fasta file**: A 'Browse...' button and a file named 'inpu_sample.fasta' with an 'Upload complete' status.
- Add new bacteriocin sequences to training (FASTA)**: A 'Browse...' button and 'No file selected'.
- Add new non-bacteriocin sequences to training (FASTA)**: A 'Browse...' button and 'No file selected'.
- Two buttons: 'Bacteriocin prediction' and 'Probability estimation' (highlighted with a white arrow).
- Two download buttons: 'Prediction results' and 'Probability results'.

On the right is the main content area with a navigation bar containing 'Welcome', 'Predicted Bacteriocin Sequences', 'Probability Scores', 'Help', and 'Data'. Below the navigation bar is a table with the following data:

Sequences	Non-bacteriocin	Bacteriocin
BAC014	0.98	0.02
BAC017	1.00	0.00
BAC023	0.98	0.02
BAC043	0.99	0.01
BAC047	0.99	0.01
BAC052	1.00	0.00
BAC058	0.99	0.01
BAC149	0.98	0.02
BAC161	1.00	0.00
BAC172	1.00	0.00
BAC177	0.99	0.01
BAC178	0.98	0.02
BAC182	0.99	0.01
WP_007235565.1	0.01	0.99
WP_007227712.1	0.01	0.99
WP_007226314.1	0.02	0.98

Fig 6: Probability estimation